INRIA

UNITÉ DE RECHERCHE
INRIA-RENNES

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P. 105
78153 Le Chesnay Cedex
France
Tél.:(1) 39 63 55 11

Rapports de Recherche

N° 1368

Programme 6
Calcul scientifique, Modélisation et
Logiciels numériques

MULTIFRAME-BASED IDENTIFICATION OF MOBILE COMPONENTS OF A SCENE WITH A MOVING CAMERA

Edouard FRANÇOIS
Patrick BOUTHEMY

Janvier 1991

RR-1368

IRISA INSTITUT DE RECHERCHE EN INFORMATIQUE
ET SYSTEMES ALEATOIRES

Campus Universitaire de Beaulieu
35042 - RENNES CEDEX
FRANCE
Téléphone : 99.36.20.00
Télex : UNIRISA 950 473F
Télécopie : 99.38.38.32

Multiframe-based identification of mobile components of a scene with a moving camera

Segmentation du mouvement dans une séquence d'images et interprétation du contenu dynamique de la scène

Edouard François† and Patrick Bouthemy‡

†*IRISA/CNRS,*
‡*IRISA/INRIA,*
Campus Universitaire de Beaulieu,
35042 Rennes Cedex, France
Tel.: 33 - 99 36 20 00 ; Fax.: 33 - 99 38 38 32
E-mail: bouthemy@irisa.fr

Publication Interne n° 564 - Décembre 1990 - 30 pages.

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
(U.R.A. 227)
UNIVERSITE DE RENNES I I.N.S.A. DE RENNES

INSTITUT NATIONAL DE RECHERCHE
EN INFORMATIQUE ET EN AUTOMATIQUE
(UNITE DE RECHERCHE DE RENNES)

Abstract

This report is concerned with the analysis and the interpretation of the dynamic content of a scene observed by a camera which can be static or mobile. The tasks involved can be for instance the detection of moving objects in a scene observed by a mobile camera, or the identification of the movements of some relevant components of the scene. This problem basically involves a motion-based segmentation step. Indeed regions corresponding to different apparent motion must be delineated before attempting to locate moving objects and to describe their motion. A new motion-based segmentation method is presented which ensures stable motion-based partitions owing to a statistical regularization approach. Besides it does not require the explicit estimation of optic flow fields. It also manages to link those partitions in time. Therefore, the motion interpretation process can be performed on more than two successive frames. The ability to follow a given coherently moving region within an interval of several images of the sequence makes the interpretation process be more robust and more comprehensive. The identification of the kinematical components of the scene relies on an intermediate layer accomplishing a generic qualitative motion labeling. Beyond the detection of mobile components of the scene observed by a moving camera, scene-related motion interpretation is expressed in terms of approaching motion, receding motion, transversal motion to the right, transversal motion to the left. This also includes quantitative parameters like time-to-collision. No 3D measurements are required. Results obtained on several real image sequences corresponding to complex outdoor situations are reported.

Résumé

Ce rapport traite de l'analyse et de l'interprétation du contenu dynamique d'une scène observée par une caméra qui peut être aussi bien statique que mobile. Il peut s'agir par exemple de détecter des objets mobiles dans une scène vue par une caméra elle-même en mouvement, ou de caractériser les mouvements dans la scène d'éléments d'intérêt. Ce problème implique fondamentalement une étape de segmentation au sens du mouvement. Il est en effet nécessaire de délimiter les régions dans l'image ayant un mouvement apparent différent avant de pouvoir localiser les objets effectivement mobiles ou de pouvoir identifier leur mouvement. Une nouvelle méthode de segmentation au sens du mouvement dans une séquence d'images a été développée. Elle ne requiert pas l'estimation explicite du champ des vitesses apparentes, et permet l'obtention de partitions stables et reliées dans le temps grâce à une approche de régularisation statistique. Ainsi l'interprétation du mouvement peut être effectuée en considérant plus de deux images successives. La capacité à suivre une région cohérente au sens du mouvement dans un intervalle de temps donné rend la phase d'interprétation plus robuste et plus complète. La caractérisation du contenu dynamique de la scène repose sur une étape intermédiaire réalisant un étiquetage qualitatif générique du mouvement. Deux tâches d'analyse de scène dynamique ont été abordées: la détection des éléments effectivement mobiles d'une scène observée par une caméra en mouvement; l'identification de trois types de mouvement utiles à la détection d'obstacles pour la navigation d'un véhicule, mouvement approchant, éloignement, mouvement latéral. L'analyse n'implique aucune mesure 3D. Un ensemble significatif d'expérimentations a été mené. Des résultats tout-à-fait probants ont été obtenus sur des séquences d'images réelles correspondant à des situations complexes de scène d'extérieur.

1. Introduction

Locating moving objects in a scene and identifying their motion is one of the main issues in dynamic scene analysis. It is relevant to a large variety of tasks such as surveillance, tracking, obstacle detection, collision course determination. If the camera is static, moving regions in the image plane necessarily correspond to moving objects in the scene. This has induced the development of methods dedicated to this situation; a first class of approaches takes into account a reference image of the static background, as in [KAR90], alternatives are also available, such as in [BOU90] where local contextual statistical model are introduced to recover moving object masks. If the camera itself undergoes motion, the solution is not as straightforward as in the previous case. This issue is addressed in [THO90]; depending on the knowledge of the camera motion, different cues are extracted from the optical flow field previously estimated to detect moving objects in the scene. That situation obviously occurs when the camera is attached to an entity whose primary function is to move, like a mobile robot, a car, or a plane. It may also be due to actions such as panning or tilting. This report is concerned with the detection of moving objects in a scene observed by a mobile camera. It also deals with the identification of their motion.

This problem basically involves a motion-based segmentation step. Indeed regions corresponding to different apparent motion must be delineated before attempting to locate moving objects and to describe their motion. It is of vital importance for the subsequent analysis to correctly achieve this early step. A new motion-based segmentation method is presented which ensures stable motion-based partitions owing to a statistical regularization approach. Besides it does not require the explicit estimation of optic flow fields. It also manages to link those partitions in time. Therefore, the motion interpretation process can be performed on more than two successive frames. The ability to follow a given coherently moving region within an interval of several images of the sequence makes the interpretation process be more robust and more comprehensive. The identification of the kinematical components of the scene relies on an intermediate layer accomplishing a generic qualitative motion labeling. The entire scheme is summarized in Fig.1. The final information level delivered depends on the task at hand. If the goal is only to detect moving objects, temporally linked maps indicate where projections of moving scene components are located in the image sequence. If we aim to know how objects move in the scene, several typical classes are handled such as approaching motion, receding motion, transversal motion to the right, transversal motion to the left. This also includes quantitative parameters like time-to-collision. An important

point is that this can be achieved without any 3D measurements. We only need to know the qualitative way the camera is moving, that is for instance panning motion or translational motion along its optical axis of view. This information is easily retrievable in any mobile rig.

The remainder of this report is organized as follows. Section 2 is concerned with the motion-based segmentation step. Section 3 deals with the symbolic kinematical labeling layer. It is based on the derivation described in [FRA90], but includes new developments corresponding to temporal extensions. Section 4 describes how interpretation of the dynamic content of the scene is handled. Results on several real image sequences corresponding to complex outdoor situations are reported.

2. Motion-based segmentation

The goal of the motion-based region segmentation step is to separate regions in the image corresponding to different motion entities. Once this partition is obtained, the kinematical analysis can be performed in each one of these regions. A motion entity is supposed to correspond to a given description level of real physical motion. In [ADI85], segmentation is performed simultaneously to the estimation of 3D motion and structure parameters from optical flow fields. Points are first gathered in elementary areas corresponding to 3D moving planar patches using a Hough transform technique. Then these areas are merged according to a hypothesize-and-verify paradigm. In [MUR87], the same description level is taken into account, i.e., planar surface patches and 3D rigid motion, but the segmentation issue is addressed as a relaxation problem based on markov random field models. In [BOU87], no 3D parameter estimation is required, since motion entities are expressed in terms of 2D velocity field models. The segmentation is achieved following a statistical approach based on likelihood ratio tests associated to a region growing procedure. Other works take profit from a multiresolution data representation, using the optical flow field or not, [BUR89], [HAR85], [PEL90].

2.1 Problem statement

The approach described hereafter makes use of 2D linear motion models, and requires no prior velocity field estimation, as in [BOU87]. The key point is that this new version is embedded in a contextual statistical framework, i.e. Markov random field (MRF) and bayesian criterion (MAP). Besides it manages to properly derive a temporal link between successive

motion-based partitions, whereas the other techniques usually do not supply such a structural information. These two characteristics jointly considered allow to get partitions stable with respect to the algorithm parametrization, and stable in time. MRF models have already been introduced in numerous low-level issues of static image analysis, and more recently of image sequence analysis, like motion detection [BOU90], motion segmentation [MUR87], optic flow estimation [HEI90], [KON89], [HUT88]. In regions where motion measurements are ill-conditioned, MRF appear as a natural way to regularize the problem at hand. Besides, due to the equivalence between MRF and Gibbs distributions, [GEM84], the modeling step reduces to the definition of a global energy function, which in turn decomposes into local potentials.

First we have to express what we consider as *observations*, and as *labels*, i.e., the motion features to be determined. Observations are directly tied to image sequence data: they are given by the spatio-temporal derivatives of the intensity function. On the other hand, label values will consist of region numbers, r, $r = 1, R$. Let S be the set of sites s where these labels are defined. S do not correspond to the pixel grid, but to a lower resolution, namely 2×2 small square windows of the original image grid. This allows to save computational time, to avoid spurious detections, and to take into account more points in the identification of motion model parameters, while still providing a sufficiently accurate segmentation. The points involved in site s will be denoted by $s_j; j = 1,4$. Let $o = \{o_s, s \in S\}$, resp. $e = \{e_s, s \in S\}$, designate the observation field, resp. the label field. The observation at site s is given by: $o_s = \{\nabla I(s_j), I_t(s_j); j = 1,4\}$, where I is the intensity function, $\nabla I = (I_x, I_y)$ its spatial gradient, I_t its temporal derivative.

Second, we have to define the criterion allowing to determine the optimal motion-based partition given the observations at hand, that is the array of spatio-temporal derivatives of the intensity function, $\nabla I, I_t$. Let us beforehand recall that this motion-based segmentation only makes sense with respect to a motion model. For the while, let us only assume that this model is parametrized by a global parameter vector θ_r, within a given region r, $r = 1, R$. Let $\theta = \{\theta_r, r = 1, R\}$. To derive the unknown label field e along with the number of regions R and the set of model parameters θ from the observed fields $(\nabla I, I_t)$, the following optimization problem has to be solved (MAP criterion):

$$\max_{e,\theta,R} \quad p(\nabla I, I_t, e, \theta, R)$$

where $p(\nabla I, I_t, e, \theta, R)$ is the joint distribution of the observed variables $(\nabla I, I_t)$ and the hidden ones e, knowing that R and $\theta = \{\theta_r, r = 1, R\}$ are additional global parameters. As previously outlined the distribution of observations and motion labels are specified using

Markov random field model whose distribution can be written as follows, [GEM84],:

$$p(\nabla I, I_t, e, \theta, R) = \frac{1}{Z}\exp - U(\nabla I, I_t, e, \theta, R)$$

where Z is a normalizing factor. The energy function U in turn decomposes into the sum of potential functions V locally defined on so-called cliques. A clique is a sub-set of sites which are mutual neighbors.

We have designed energy function U as the combination of two terms:

$$U(\nabla I, I_t, e, \theta, R) = U_1(e) + U_2(\nabla I, I_t, e, \theta, R). \quad (1)$$

U_1 contains the expected generic properties of the resulting label field, (regularization effect), while U_2 expresses the adequacy between observations and labels.

2.2 Modeling step

Prior model, energy term U_1

Here we consider a second-order neighborhood system, and cliques used are binary ones, denoted $c =< s, t >$, (see Fig.2). Let C be the set of all binary cliques in the image. The potential function is defined as follows:

$$\begin{aligned} V(e_s, e_t) &= \begin{cases} 0 & \text{if } e_s = e_t \\ \alpha & \text{otherwise} \end{cases} \\ &= \alpha\ (1 - \delta_{e_s = e_t}) \end{aligned}$$

where δ designates the Kronecker symbol and α is a positive predetermined parameter. These potentials have been defined in such a way as to favouring homogeneity of the label field, knowing that a positive value of V discourages the corresponding configuration. Other kinds of potential functions have been tested, for instance based on parameter vectors θ_r. In fact this simplest one has been proved the most efficient one. Accordingly the first energy term is given by:

$$U_1(e) = \sum_{c \in C} \alpha\ (1 - \delta_{e_s = e_t}) \quad (2)$$

Observation-label interaction model, energy term U_2

We have now to express the interaction model between observations and labels through energy term U_2. The starting idea is that we have to compare the velocity vectors corresponding to the considered motion model to the true velocity vectors. A region can be stated as coherent with respect to motion cue if its motion content can be described through one model parametrization. Conversely a point can belong to region r if its true velocity vector

is close to the one induced at this point by the motion model parametrized by vector θ_r. Let us denote the model-related velocity vector field by $\vec{\omega}_{\theta_r}$. Formally, this leads to take into account the difference: $\vec{\omega} - \vec{\omega}_{\theta_r}$. If the velocity vector field is not available, this quantity can not be directly evaluated. However the spatio-temporal derivatives of the intensity function are related to the apparent velocity vector by the well-known image flow constraint equation, [SCH86]:

$$\nabla I.\vec{\omega} + I_t = 0$$

Therefore let us consider the following quantity at each point s_j:

$$\epsilon(s_j) = \nabla I(s_j)(\vec{\omega_{\theta_r}}(s_j) - \vec{\omega}(s_j))$$

Using the above mentioned image flow constraint, this can be written as follows:

$$\epsilon(s_j) = \nabla I(s_j).\vec{\omega_{\theta_r}}(s_j) + I_t(s_j)$$

If correctly parametrized, the variables ϵ are assumed to follow a zero-mean gaussian law of variance σ^2. Moreover they are supposed to be independent. In practice, 2D linear models are used:

$$\theta_r = (a, b, \alpha, \beta, \gamma, \delta)$$

and at point (x, y), the velocity vector generated by the linear motion model parametrized by θ_r is given by:

$$\vec{\omega}_{\theta_r}(x, y) = \begin{pmatrix} a + \alpha.\Delta x + \gamma.\Delta y \\ b + \beta.\Delta x + \delta.\Delta y \end{pmatrix}$$

where Δx, and Δy designate the differential shift with respect to a reference point, which will be the region gravity centre. The choice of such a linear model represents a good trade-off between model complexity and model efficiency.

The energy term U_2 can now be expressed as follows:

$$U_2(\nabla I, I_t, e, \theta, R) = \sum_{s \in S} \left(\frac{1}{2\sigma^2} \sum_{j=1}^{4} (\nabla I(s_j).\vec{\omega}_{\theta_{e_s}}(s_j) + I_t(s_j))^2 \right) \tag{3}$$

Finally, maximizing the joint distribution (??) is equivalent to minimizing the following energy function U:

$$U = U_1 + U_2 = \alpha \sum_{c \in C} (1 - \delta_{e_s = e_t}) + \sum_{s \in S} \left(\frac{1}{2\sigma^2} \sum_{j=1}^{4} \left(\nabla I(s_j).\vec{\omega}_{\theta_{e_s}}(s_j) + I_t(s_j) \right)^2 \right) \tag{4}$$

Let us point out that the same framework can take directly into account as observation the optic flow field if available. We only need to define variable ϵ by $\epsilon(s_j) = \|\vec{\omega}_{\theta_r}(s_j) - \vec{\omega}(s_j)\|$.

2.3 Relaxation scheme

The minimization of energy function U is carried out according to a deterministic relaxation, of the ICM kind. The initial state of the variables to be identified at time t is derived from the optimal values obtained at time $t-1$ as explained in subsection 2.4 (apart from the case corresponding to the very beginning of the sequence). The estimation of label field e and the identification of global parameters θ are performed in an alternate iterative way. We mean that we do not continuously update the values of parameters θ_r while modifying the estimate of label field e. Briefly, we perform a number of minimization iterations concerning only label field e while required, then within each currently delimited region r we estimate the optimal parameter values $\hat{\theta}_r$; then we start again the estimation of label field e, and so on. The main reason is that parameters θ_r are global ones, whereas label field e is very locally updated.

The estimation of the parameter vector θ_r for each region r is achieved by minimizing energy term U_2 while freezing the other variables, i.e. the label field e. This leads to a least-square estimation of the parameters θ_r within each region r. Before explaining how the minimization of the energy function U with respect to e is pursued, let us comment how we tackle the problem of deciding what is the right number of regions $\hat{R}$ to be considered. This number is usually predetermined in markovian approaches dealing with image segmentation issues. But it is not obvious to a priori define the number of different motion entities in an image sequence. Therefore we take $\hat{R}$ equal to R_0 at the very beginning of the processed image sequence and we update it as explained at the end of this subsection. The value R_0 depends on the choice of the initialization procedure. If the very initial label field consists of one region for the entire image, $R_0 = 1$; if we use the technique described in [BOU87] to get an initial partition based on the two first images of the sequence, R_0 is equal to the number of regions found in this initial partition.

The estimation of label field e only involves very local computations. Sites are visited one by one. As a matter of fact, we have a list of sites to be visited at iteration i, which has been built up during iteration $i-1$. At a given site s, all the possible discrete label values from 1 to $\hat{R}$ should be tested. This could induce a tremendous amount of computations depending on the current value of $\hat{R}$. In fact this can be reasonably avoided in our case. Indeed the purpose here is not to classify but to segment into connected components. The labels likely to be candidate at site s are those which are already present in its neighborhood, including the one currently assigned to it. Nevertheless a supplementary label ρ must always be considered. In practice, first the label value $\hat{r}_i$ among the ones existing in the neighborhood of s apart

from ρ is determined which minimizes the local energy defined by:

$$\Delta U_s(r) = \sum_{j=1}^{4} (\nabla I(s_j).\vec{\omega}_{\theta_r}(s_j) + I_t(s_j))^2 + A \sum_{<s,t>} (1 - \delta_{e_t=r})$$

where $A = 2\sigma^2\alpha$. Then the local energy is evaluated for the label ρ:

$$\Delta U_s(\rho) = \Phi + A \sum_{<s,t>} (1 - \delta_{e_t=\rho})$$

where Φ is positive parameter, which can be explicitly determined using tables associated to χ^2 statistical law. Indeed the restriction of U_2 to the the considered neighborhood, can be assumed to follow a χ^2 distribution with four degrees of freedom (since the 2×2 square windows contain four points), if the site is assigned the right label. If we want the probability to choose $\hat{r}_i$ be greater for instance than 0.9999, when it is the right one, we must take $\Phi = 25$. The label value finally selected is the one between $\hat{r}_i$ and ρ which minimizes the local energy ΔU_s.

After a certain quantity of iterations, we assign numbers ranging from $\hat{R} + 1$ to R' to the areas of sufficient size labeled by ρ. The process is then rerun with $\hat{R}$ set to R'. This is iterated until convergence is reached (that is when the number of points whose label has been changed is below a given threshold).

The use of Markovian models and of the associated relaxation process brings two major interests, in comparison with the region-growing algorithm described in [BOU87]. First, segmentation is done only using local and regular computations, whereas the region-growing algorithm induces a less workable implementation. Second, this statistical regularization method ensures more stable results, than any region-growing procedure, usually sensitive to the choice of input parameter values and the region merging order.

2.4 Temporal link of the successive partitions

When the process is completed at time t, the motion-based segmentation map at time t (denoted $\wp_t$) is available. Moreover for every region of this map, linear motion model parameters have been estimated, that is $\hat{\theta} = \{\hat{\theta}_r, r = 1, \hat{R}\}$. $\hat{\theta}_r$ is the estimated parameter vector of region r. It is straightforward to compute the corresponding model velocity field from time t to time $t + 1$, $\vec{\omega}_{\hat{\theta}_r,t}$. Let us consider a point $p(x, y)$ which belongs to region r. At point p, the model-related velocity vector is determined by:

$$\vec{\omega}_{\hat{\theta}_r,t}(p) = \begin{pmatrix} \hat{a}_r + \hat{\alpha}_r \Delta x + \hat{\gamma}_r \Delta y \\ \\ \hat{b}_r + \hat{\beta}_r \Delta x + \hat{\delta}_r \Delta y \end{pmatrix}$$

Since this model parametrization is the optimal one with respect to the segmentation stage, it seems reasonable to use this model-related velocity field as follows. Using the segmentation map $\wp_t$ and velocity field $\vec{\omega}_{\hat{\theta}_{r,t}}$ at time t, we can build up a prediction $\breve{\wp}_{t+1}$ of the segmentation map at time $t+1$. This is performed by projecting $\wp_t$ in the direction of motion given by $\vec{\omega}_{\hat{\theta}_{r,t}}$. It may happen that at some points, no label is predicted (because no velocity vector points toward this point), or on the contrary more than one label are predicted (several velocity vectors point toward this point). An interpolation step is achieved in order to correctly label these points. Thus an initial motion-based segmentation map is avalaible at time $t+1$. Moreover this process enables to link in time the different regions.

To determine if each region of $\breve{\wp}_{t+1}$ map keeps on undergoing a similar motion as the one at time t of the region to which it is linked through this projection-interpolation process, a first relaxation stage is performed taking into account motion parameters of time t, $\hat{\theta}_{r,t}$. Hence, if the underlying 3D motion of a region has significatively changed from time t to time $t+1$, the same holds for apparent motion parameters of this region, accordingly this region will be labeled by ρ. This first stage ensures that each region corresponds to a scene component with a coherent 3D motion along the sequence (or a part of it). This is important for the further interpretation step which makes use of several successive partition maps. Once this first relaxation stage is achieved, the motion model parameters at time $t+1$ are effectively estimated from this first updated version of the motion-based segmentation map $\wp_{t+1}$. The complete relaxation process is then run as described above.

When the segmentation of a subset of successive frames is achieved, a global sequence segmentation is available. The same label is assigned to each motion-based coherent region along the sequence. Indeed, we get several "motion tubes", each one corresponding to one single motion entity. Afterwards, each entity could be described with global parameters such as 3D global motion and structure parameters.

2.5 Results

The first considered sequence contains two cars manoeuvring in a park (Fig. 5). The first car in the foreground is rotating and moving toward the camera. The second one in the background is translating toward the camera. It partially hides a third car which is static. The camera pans the scene from right to left. The image background consists of trees, some of them being stirred by the wind (in particular in the centre of the image plane). Initialization has been performed using the method presented in [BOU87]. Segmentation has been

achieved with $A = 25$ and $\Phi = 25$ Results are shown in Fig. 6. The obtained motion-based partition maps are quite stable in time and correctly linked from one image to the next. It can also be notice that the linear motion model level is quite relevant and delivers motion entities of practical use in the context of dynamic scene analysis.

The second sequence (Fig. 7) corresponds to a slow camera motion, which is translating along its optical axis, in an urban context. The images include on the right a traffic light, and in the middle a car which is transversally translating to the right. The initial partition at the very beginning of the sequence is merely taken as a single region corresponding to the whole image. The value of Φ is the same (i.e., equal to 25), though A is set to 15. In fact, this sequence is less complex and noisy than the previous one. The parameter A, which is proportional to the variance of the noise, must be greater for the first sequence than for the second one. Results are again satisfactory (Fig. 8). A part in the middle of the car is merged with the image background. As a matter of fact the intensity function of this area is quite uniform. Indeed it is very hard to see locally any motion or intensity boundary. That is the reason why this part is merged to the background region. The apparent motion of the background region is due to the camera movement. The same holds of course for the region corresponding to the traffic light; but since there is a great difference in depth between this object and the rest of the scene objects, it is normal that there is a discrepancy in the perceived motion. Again, it is interesting to point out that this motion model level naturally reflects both 3D motion and depth information, which is very important for instance in case of an obstacle detection task.

It can be seen with these two examples that the proposed motion segmentation method yields good results in different situations. Others sequences have been treated, and satisfactory results have been again obtained. This algorithm is not cpu time consuming. A segmentation process of a sequence of 20 images, of size 224×224 pixels (whose output is 19 successive motion-based partition maps) takes about 80 seconds using a SUN-4 workstation, that is 4 seconds per image. The main problem is the setting of the contextual parameter A. It is necessary to take into account the complexity of the scene, the magnitude of apparent motions and the intrinsic noise contained in the images. Roughly, if the scene is simple, with small motions and few noise, this parameter must be fixed at a low value. On the contrary if motions are important, it can be noticed that the underlying photometric model leading to the optic flow constraint equation becomes inexact. In this case, A must be set to a great value. However, for the whole set of sequences that we have processed, only two values for A have to be used (15 and 25).

3. Symbolic labeling

Before handling the scene-related motion interpretation step, which in fact depends on the considered application, an intermediate layer is considered. This process provides a generic qualitative kinematical labeling on which any further interpretation will rely on. Its purpose is to describe any object movements by its motion type (e.g., translation or rotation) and by its trajectory type (e.g., parallel to the optical axis or perpendicular to this it). The basic idea of this approach is that the geometry of the 2D velocity vector field in the image contains information about the 3D motion, which can be significant and sufficient enough for interpretation. Besides a linearized version of the velocity field is considered, which is coherent with the motion segmentation step. The method we use to derive this qualitative labeling is described in [FRA90]. In the following we will only briefly recall the main aspects of this method. We will focus on important new temporal extensions which are of key interest for interpretation.

3.1 2D-3D relations

The first point is to introduce relevant terms, describing specific forms of the vector field, which could be easily and naturally interpreted. Some studies and experiments allow to conclude that a first order development of a vector field can be sufficient to interpretation; [MAI89], [FRA90], [SUB90]. The first order development of the 2D velocity vector field in the image is then considered, and four terms which well describe the geometry of the velocity field are derived. These terms are called divergence, rotational and hyperbolic terms. They will be respectively denoted by $div, rot, hyp1$ and $hyp2$. They correspond to fields which belong to four orthogonal supplementary subspaces. This is exemplified in Fig.3.

Any vector field can be approximated by a combination of a divergent field, a rotational field, and two hyperbolic fields. In case of 2D velocity vector fields, projection in the image plane of 3D velocity vector fields, a particular kind of motion in the scene implies a particular combination of these basic fields. Indeed, a divergent velocity vector field is the result of an axial motion, i.e. along the optical axis of the sensor. A rotational field appears as the result of a rotation around this axis. Let us now briefly outline the link between 3D motion parameters and each of the four terms above mentioned. Relations of this kind are also derived in [SUB90] but in a slightly different way.

Let us consider the coordinate system given in Fig.4. Let $\vec{T} = (U, V, W)^T$ and $\vec{\Omega} =$

$(A, B, C)^T$ respectively denote the instantaneous translation vector and rotation vector representing any 3D rigid motion [LON80]. Let Z_0, γ_1, γ_2 be the first order parameters of the considered object surface patch (depth and orientation parameters). Using the well-known relation (issued from the kinematical screw) between the 3D velocity vector $\vec{V} = (\dot{X}, \dot{Y}, \dot{Z})^T$ of a point $P = (X, Y, Z)$ on the object surface, and the instantatneous translation and rotation vectors attached to this object, that is: $\vec{V} = \vec{T} + \vec{\Omega} \wedge \vec{OP}$; using relations between $\vec{V}$ and $\vec{\omega}$, the 2D velocity vector of point p, projection in the image of point P; and considering a linearized version of the 2D velocity field, and the surface function Z

$$Z = Z_0 + \gamma_1 X + \gamma_2 Y \tag{5}$$

it can be shown, [FRA90], that

$$\begin{aligned} div &= -2\frac{W}{Z_0} - \gamma_1 \frac{U}{Z_0} - \gamma_2 \frac{V}{Z_0} & rot &= 2C - \gamma_1 \frac{V}{Z_0} + \gamma_2 \frac{U}{Z_0} \\ hyp1 &= -\gamma_1 \frac{U}{Z_0} + \gamma_2 \frac{V}{Z_0} & hyp2 &= -\gamma_1 \frac{V}{Z_0} - \gamma_2 \frac{U}{Z_0} \end{aligned} \tag{6}$$

The qualitative analysis of the 3D motion will rely on these relations (6).

3.2 Qualitative description

As far as we are concerned with a qualitative interpretation of motion, what is important is the comparison of these terms to zero. Accordingly, to each quantitative term $div, rot, hyp1, hyp2$ is associated a qualitative (boolean) variable $V_{div}, V_{rot}, V_{hyp1}, V_{hyp2}$, equal to 0 if its quantitative value is non significant, and respectively $D, R, H1, H2$ otherwise.

From the relations (4), it is clear that several typical dynamic situations can be easily expressed. For instance, in the $(V_{div}, V_{rot}, V_{hyp1}, V_{hyp2})$ basis, the $(D, 0, 0, 0)$ symbol association is the qualitative description of a motion along the optical axis. In [NEL88] an obstacle detection task is handled according to a qualitative way of reasoning and is based on such a a divergence cue. $(0, R, 0, 0)$ symbol association describes a rotation around this axis, and so on. In the remainder, we will also call a symbol association a label (the word "label" is used here with a different meaning than in the section devoted to the motion-based segmentation step). The great interest of such a symbolic description level will be exemplified in the next section.

It is worth stressing that the motion description is only based on the linear terms of the velocity field. These terms do not depend on the reference point in the image. Therefore the position of the optical center is not needed, and the FOE must not be estimated. Moreover,

as the goal is to obtain a qualitative description, and not the precise quantitative estimation of 3D parameters, a complete camera calibration is not required.

3.3 Labeling process

Once the motion-based image partition has been obtained, motion of each region can be qualitatively labeled. The key point is to properly achieve the numerical-to-symbolic step: that is deriving symbols from numerical data. Given an area in the image, the most convenient label must be determined, i.e. the best fit to the underlying 3D object motion. The simple comparison of the magnitudes (or function of the magnitudes) of the quantitative terms $div, rot, hyp1, hyp2$ to a threshold, as initialy described in [THO84], remains very difficult and tricky. On one hand, the threshold choice would be very dependent on the context and goal of the analysis. On the other hand, a proper noise modeling with such a procedure cannot be introduced, whereas the considered observations (related to optical flow) are actually noisy. Moreover models that we deal with are not perfectly exact (first order). That is the reason why a statistical approach is used to cope with this problem. The right state of each qualitative variable V_{div}, V_{rot}, V_{hyp1}, V_{hyp2} must be determined. To this end a method based on statistical tests has been designed, namely likelihood tests which are known to be very powerful, [FRA90]. V_{div} state is set by testing hypothesis $\underline{H_0}$: (D, R, H_1, H_2) against hypothesis $\underline{H_1}$: $(0, R, H_1, H_2)$. This formalizes that the state of V_{div} is determined while letting the three other variables free. Practically, first the optimal parameter vector $\hat{\psi}_0 = (\hat{div}, \hat{rot}, \hat{hyp1}, \hat{hyp2})$ for the (D, R, H_1, H_2) hypothesis, i.e. which maximizes the likelihood function $f(\psi_0)$, is estimated. The likelihood function is the joint density distribution, within the considered region r, of variables similar to the ones considered in the motion-based segmentation step:

$$\epsilon_\psi(x, y) = \nabla I(x, y).\vec{\omega}_\psi(x, y) + I_t(x, y)$$

which are again supposed to be independant gaussian zero-mean variables. (The same remark holds concerning the choice of the form of this variable if the velocity field is available). $\hat{\psi}_1 = (0, \hat{rot}, \hat{hyp1}, \hat{hyp2})$ for the $(0, R, H_1, H_2)$ hypothesis is similarily determined. Then the likelihood ratio $L(V_{div}) = \log\left(f(\hat{\psi}_1)/f(\hat{\psi}_0)\right)$ is compared to a threshold λ. If the ratio is lower than this preset threshold, the V_{div} symbol is equal to D, 0 otherwise:

$$L(V_{div}) \underset{H_0\ (D)}{\overset{H_1\ (0)}{\gtrless}} \lambda$$

Depending on the application, the state of each symbol is determined through a predefined decision tree. Once each symbol state is determined, the optimal label is defined as their association. It appears, [FRA90], that only five associations among the sixteen possible configurations have a physical meaning when dealing with rigid motion, that is $(0,0,0,0)$; $(D,0,0,0)$; $(0,R,0,0)$; $(D,R,0,0)$; $(D,R,H1,H2)$. The first one correspond to either static part, a pan or tilt motion, or a lateral translation of planar surface parallel to the image plane. The two next ones have already been commented. The fourth one is a combination of the two last ones. The fifth one comprises all more complex combinations.

3.4 Temporal extensions

The four terms introduced above supply an instantaneous description of the motion, that is, related to the geometry of the velocity field between two frames. In practice, some instabilities in successive labeling maps may appear. Then it is interesting to consider several successive frames as a global set of observations, and to perform the labeling process globally on these successive images. This is reachable since we have defined in Section 2 a motion-based segmentation method which yields temporally-linked partition maps. The advantages of such an approach are two-fold. First since a larger temporal interval is taken into account, robustness is improved. Second this also enables to make the interpretation level be more complete. The consideration of a temporal dimension in the 3D structure-and-motion-from-2D-motion issue has already been proved attractive, [SUB89].

Nevertheless, even if the 3D motion of an analyzed object is constant in time, the four description terms will change in time. Hence, temporal quantities need to be introduced to express the temporal dimension of the problem at hand: namely the derivatives with respect to time of the four terms defined above, $\dot{div}$,$\dot{rot}$,$\dot{hyp1}$,$\dot{hyp2}$. These new parameters can be used for interpretation. Specific symbol associations again correspond to specific kinematical configurations. Even if motion interpretation is still performed considering only the four terms $div, rot, hyp1, hyp2$, the temporal parameters $\dot{div}, \dot{rot}, \dot{hyp1}, \dot{hyp2}$ must be taken into account in the parameter estimation step.

We consider a region, whose 3D motion parameters (U,V,W) and (A,B,C) are constant within the time interval corresponding to the subset of successive frames to be considered. The movement in space of an object usually implies changes in the structure parameters Z_0, γ_1 and γ_2.

The introduction of the temporal parameters is done through a linear development in time of the linearized version of the instantaneous velocity field. As previously defined, we

have:

$$\vec{\omega}_{\theta_r}(x,y) = \begin{pmatrix} a \\ b \end{pmatrix} + M \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$$

where $(a,b)^T$ is the constant part of the decomposition, and M is a matrix linearly depends on the four terms $div, rot, hyp1, hyp2$, which can be written as follows:

$$M = \frac{1}{2}\, div\, [D] + \frac{1}{2}\, rot\, [R] + \frac{1}{2}\, hyp1\, [H1] + \frac{1}{2}\, hyp2\, [H2]$$

with

$$[D] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, [R] = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, [H_1] = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, [H_2] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Adding a temporal dimension to this development leads to:

$$\vec{\omega}_{\theta_r}(x,y) = \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} \dot{a} \\ \dot{b} \end{pmatrix} \Delta t + \left(M + \dot{M}\Delta t\right) \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$$

It can be easily shown that:

$$\dot{M} = \tfrac{1}{2}\, \dot{div}\, .\Delta t\, [D] + \tfrac{1}{2}\, \dot{rot}\, .\Delta t\, [R] + \tfrac{1}{2}\, \dot{hyp1}\, .\Delta t\, [H1] + \tfrac{1}{2}\, \dot{hyp2}\, .\Delta t\, [H2]$$

knowing that

$$\begin{cases} \dot{div} = \dot{\alpha} + \dot{\delta} & \dot{hyp1} = \dot{\alpha} - \dot{\delta} \\ \dot{rot} = \dot{\beta} - \dot{\gamma} & \dot{hyp1} = \dot{\beta} + \dot{\gamma} \end{cases}$$

Some elementary mathematical developments allow to obtain the following relations:

$$\begin{cases} div = -\gamma_1 \frac{U}{Z_0} - \gamma_2 \frac{V}{Z_0} - 2\frac{W}{Z_0} & \dot{div} = (-\dot{\gamma_1}\frac{U}{Z_0} - \dot{\gamma_2}\frac{V}{Z_0}) - \frac{\dot{Z_0}}{Z_0}\, div \\ rot = -\gamma_1 \frac{V}{Z_0} + \gamma_2 \frac{U}{Z_0} + 2C & \dot{rot} = (-\dot{\gamma_1}\frac{V}{Z_0} + \dot{\gamma_2}\frac{U}{Z_0}) - \frac{\dot{Z_0}}{Z_0}(-\gamma_1 \frac{V}{Z_0} + \gamma_2 \frac{U}{Z_0}) \\ hyp1 = -\gamma_1 \frac{U}{Z_0} + \gamma_2 \frac{V}{Z_0} & \dot{hyp1} = (-\dot{\gamma_1}\frac{U}{Z_0} + \dot{\gamma_2}\frac{V}{Z_0}) - \frac{\dot{Z_0}}{Z_0}\, hyp1 \\ hyp2 = -\gamma_1 \frac{V}{Z_0} - \gamma_2 \frac{U}{Z_0} & \dot{hyp2} = (-\dot{\gamma_1}\frac{V}{Z_0} - \dot{\gamma_2}\frac{U}{Z_0}) - \frac{\dot{Z_0}}{Z_0}\, hyp2 \end{cases}$$

The last point is to express $\dot{Z}_0, \dot{\gamma_1}$ and $\dot{\gamma_2}$. To this end, relation (5) is temporally derived, then $\dot{X}, \dot{Y}, \dot{Z}$ are substituted for their expressions in the 3D velocity vector $\vec{V}$, and Z for its expression given in relation (5). Terms of similar order are grouped, and the following relations are finally derived:

$$\begin{cases} \dot{Z}_0 = W(1 - \gamma_1 B + \gamma_2 A) - \gamma_1 U - \gamma_2 V \\ \dot{\gamma_1} = -B(1 - \gamma_1 B + \gamma_2 A) - \gamma_2 C \\ \dot{\gamma_2} = A(1 - \gamma_1 B + \gamma_2 A) + \gamma_1 C \end{cases}$$

It is now possible to relate temporal terms $\dot{div}, \dot{rot}, \dot{hyp1}, \dot{hyp2}$ to 3D motion configurations. A reasoning similar to the one done for the spatial terms $div, rot, hyp1, hyp2$ in section 3.2 allows to establish new more complete labels corresponding to specific kinematical configurations. Indeed the five labels of interest extend as follows: $(0,0,0,0)$ into $(0,0,0,0,0,0,0,0)$, $(D,0,0,0)$ into $(D,0,0,0,d,0,0,0)$, $(0,R,0,0)$ into $(0,R,0,0,0,0,0,0)$, $(D,R,0,0)$ into $(D,R,0,0,d,0,0,0)$, and $(D,R,H1,H2)$ into $(D,R,H1,H2,d,r,h1,h2)$, where $d,r,h1,h2$ are the symbols corresponding to significant values of resp. $\dot{div}$, $\dot{rot}$, $\dot{hyp1}$, $\dot{hyp2}$. These new labels can be used for motion interpretation in the same way than the five "instantaneous" labels. They can induce a more robust in the labeling step. Supplementary relations can also be found between those eight terms, leading to a more precise motion description. In particular, it becomes possible to subdivide the large class corresponding to the label $(D,R,H1,H2)$ (transversal translation of a surface non parallel to the image plane, combined with any other motions). Within this class indeed specific new relations can be explicitly written:

- if the motion is a pure transversal translation of a non parallel surface:

$$\frac{\dot{div}}{div} = \frac{\dot{rot}}{rot} = \frac{\dot{hyp1}}{hyp1} = \frac{\dot{hyp2}}{hyp2} = -div \tag{7}$$

- if this translation is combined with a translation along the optical axis:

$$\frac{\dot{div}}{div} = \frac{\dot{rot}}{rot} = \frac{\dot{hyp1}}{hyp1} = \frac{\dot{hyp2}}{hyp2} \neq -div \tag{8}$$

- if this translation is combined with a rotation around the optical axis:

$$\frac{\dot{div} + div^2}{rot} = -\frac{\dot{rot} + rot^2}{div} = -\frac{\dot{hyp1} + hyp1^2}{hyp2} = \frac{\dot{hyp2} + hyp2^2}{hyp1} \tag{9}$$

If a motion is labeled $(D,R,H1,H2,d,r,h1,h2)$, but none of relations (7), (8), (9) is verified, we can conclude that the motion is a complex combination of a transversal translation with others motions.

4. Scene-related motion interpretation

The symbolic kinematical labeling is a very general intermediate layer which can be of practical use in various situations to achieve scene-related motion interpretation. The labels defined above can be used in the interpretation step in different ways depending on the goal

of the application at hand. This will be illustrated by two examples of dynamic scene analysis tasks relevant to significant applications. The first one is concerned with the detection of moving objects in the scene knowing the type of camera motion. This situation is a current one in numerous surveillance applications. The second one copes with an obstacle detection issue in a car driving assistance application. We assume that the camera is translating along its optical axis.

4.1 Example 1: Locating moving objects

This first example corresponds to the following camera motion, pan motion. This means that the camera is rotating around the vertical axis (Y), that is using the notations introduced in Section 3, $B \neq 0$. The goal is to locate moving objects in the scene. Obviously the interpretation will reduce into two classes: moving objects and static objects (in the absolute coordinate system). In this situation, to any static object in the scene should be associated an apparent motion corresponding to the label (0,0,0,0), according to relations (), since in this case only parameter B is non zero. Therefore if a region in the image is assigned any another label $((D,0,0,0),(0,R,0,0),(D,R,0,0)$ or $(D,R,H1,H2))$, this implies that the corresponding object in the scene is itself moving. Hence it is quite straightforward to detect mobile objects, using only qualitative labels. Of course situations correrresponding to other camera movements could be similarily handled.

Results

As previously outlined labeling is practically achieved using a small set of successive images from the sequence. Here we take five successive images. The analyzed sequence is the sequence shown in Fig. 5). Three regions correspond to mobile entities in the scene: the car in the foreground (turning toward the camera), another one in the background (translating toward the camera), and the region in the middle of the images corresponding to trees significantly perceived as stirred by the wind. Generic symbolic motion labeling has been achieved with a threshold value in the likelihood test of $\lambda = 0.01$. Regions textured by vertical lines correspond to moving components in the scene, whereas white regions are static parts of the scene. The interpretation process gives good results, since only one region is misclassified. It is located on the top of the image, and mainly corresponds to a piece of sky. This area is nearly uniform, hence no motion information is truly available inside it. This can explain that the parameter estimation is certainly poor, and infers a wrong motion characterization.

On the other hand, mobile components of the scene are quite well located all along the sequence, although complex (and some non rigid) apparent motions are observed.

4.2 Example 2: Recognizing 3D kinematical behaviours

Scene-related kinematical description can also be more complete. Indeed, many applications require informations about object motion in the scene. A typical example is an obstacle detection task for navigation purpose. Indeed we are concerned with an application dealing with car driving assistance. We aim to supply information about kinematical behaviour of the other scene components. We consider a car equipped with a camera looking ahead. The car is supposed to move parallel to the optical axis. The problem is here to provide alarms, when an "object" is potentially dangerous, e.g. if it is moving toward the camera or may have a collision course. Three interpretation classes are defined: objects moving toward the camera (class 1), objects moving away from the camera (class 2), and objects with a transversal motion (class 3). In this situation, three labels are of real interest: $(D, 0, 0, 0)$ which corresponds to a relative translation along the optical axis, (classes 1 and 2), $(D, R, H1, H2)$ which corresponds to entities undergoing motion with a significant lateral component, (class 3), $(0, 0, 0, 0)$ corresponding to static part when the car stops, we have also in this case $(a, b) = (0, 0)$, otherwise it is a planar surface whose orientation and motion are parallel to the image plane (class 3). Classes 1 and 2 are further discriminated depending on the sign of the div term. Quantitative measurements are also available like time-to-collision given by $1/div$. If the object is labeled by $(0, 0, 0, 0)$, the estimation of the constant term a can be used to know the direction of the transversal translation (left if a is negative, right if a is positive). Finally, if the object is labeled by $(D, R, H1, H2)$, this information can only be used if the position of the optical center is roughly known.

Results

The labeling step is again achieved considering several successive linked partitions $\hat{\wp}_t$. For display convenience, three textures are used. If an object is receding, the corresponding region is textured with lines oriented to the left, if it is approaching, it is textured with full grey, and if it is transversally translating, it is textured with lines oriented to the right. In this last case a small arrow indicates the direction of the translation.

Experiments with several real sequences have been carried out. We only report here two of them. In the first one (Fig. 11), the camera is static. A pedestrian on the left, and a cyclist in the middle are transversally translating to the left. On the right, a second pedes-

trian is moving away from the camera. The interpretation process yields the right label for each one of these mobile scene components (Fig. 12). In the second sequence (Fig. 14), the camera is attached to a car moving on the right lane of the road in a direction parallel to the camera axis. The own car is crossing another car which is also translating toward the camera along a direction parallel to the optical axis. On the right, there is a pole which is static. Both objects should be considered as potentially dangerous, since their relative motion with respect to the camera is an approaching one. Effectively, they are labeled by the right label "approaching" (fig 13). In both examples, the threshold value λ in all the likelihood tests involved in this labeling process is equal to 0.01. This proves that to set this threshold to a relevant value is not a critical matter at all. Besides it is worth noticing that the interpretation classes are correctly recognized all along the sequence. Since the link between regions in successive partition maps are retrievable, some kind of "intelligent" tracking could be considered.

5. Conclusion

This report has described a complete scheme for analyzing the dynamical content of a scene observed by a moving camera. It takes into account as input a monocular image sequence and involves three main steps: 1) a motion-based segmentation module allowing to track regions along the sequence; 2) a spatio-temporal generic symbolic motion labeling module; 3) a scene-related motion interpretation module which is straightforwardly derived from the previous one according to the application at hand. Several experiments have been reported dealing with real sequence depicting complex outdoor situations, and corresponding to relevant dynamic scene analysis tasks. The interests of this study are three-fold. First this method does not require neither explicit 3D measurements nor the estimation of 2D velocity field. It mainly relies on the spatio-temporal variations of the intensity function while making use of motion models at an appropriate level (2D linear models which are a good trade-off between computation complexity and description efficiency). Second since we deal with not perfectly exact models and noisy measurements, statistical criteria have been considered for the two first basic steps of this interpretation scheme. Third the temporal dimension of the problem is properly taken into account. The ability to supply temporally-linked motion-based segmentation maps enables the motion interpretation step to rely on more than two successive frames, which makes it more robust and comprehensive.

Acknowledgments

This work is supported by MRT (French Ministery of Research and Technology) in the context of the EUREKA european project PROMETHEUS, under PSA-contract VY/85241753/14/Z10, and by Région Bretagne (Brittany County Council) under contribution to student grant.

We thank Dr Enkelman for providing the image sequences of Fig.7, Fig.11 and Fig.14.

References

[ADI85] ADIV G. Determining three-dimensional motion and structure from optical flow generated by several moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 7:pp 384–401, Jul. 1985.

[BOU87] BOUTHEMY P., SANTILLANA RIVERO J. A hierarchical likelihood approach for region segmentation according to motion-based criteria. *Proc. 1st International Conference on Computer Vision, London*, pp 463–467, 1987.

[BOU90] BOUTHEMY P., LALANDE P. Detection and tracking of moving objects based on statistical regularization method in space and time. *Proc. First European Conference on Computer Vision, Antibes, France*, pp 307–311, April 1990.

[BUR89] BURT P.J., BERGEN J.R., HINGORANI R., KOLCZYNSKI R., LEE W.A., LEUNG A., LUBIN J., SHVAYTSER H. Object tracking with a moving camera. *Workshop on Visual Motion*, pp 2–12, March 1989.

[FRA90] FRANCOIS E., BOUTHEMY P. The derivation of qualitative information in motion analysis. *Image and Vision Computing Journal*, Nov. 1990.

[GEM84] GEMAN S., GEMAN D. Stochastic relaxation, Gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.6, No.6:pp 721–741, Nov. 1984.

[HAR85] HARTLEY R. Segmentation of optical flow fields by pyramid linking. *Pattern Recognition Letters*, Vol.3:pp 253–262, Jul. 1985.

[HEI90] HEITZ F., BOUTHEMY P. , BOUTHEMY P. multimodal motion estimation and segmentation using markov random fields. *Proc. 10th IEEE Conference on Pattern Recognition, Atlantic City*, pp 378–383, June 1990.

[HUT88] HUTCHINSON J., KOCH C., LUO J., MEAD C. Computing motion using analog and binary resistive networks. *Computer*, Vol.21: pages 52–63, March 1988.

[KAR90] KARMANN K.-P., V.BRANDT A., GERL R. Moving object segmentation based on adaptative reference images. *Proc. Conference Eusipco, Barcelone*, pp 951–954, Sept. 1990.

[KON89] KONRAD J., DUBOIS E. Bayesian estimation of discontinuous motion in images using simulated annealing. *Proc. Vision Interface, London, Ontario, Canada*, June 1989.

[LON80] LONGUET-HIGGINS H.C., PRAZDNY K. The interpretation of a moving retinal image. *Proc. Roy. Soc. Lond.*, B-208:pp 385–397, April 1980.

[MAI89] MAILLOUX G.E., LANGLOIS F., SIMARD P.Y., BERTRAND M. Restoration of the velocity field of the heart from two-dimensional echocardiograms. *IEEE Transactions on Medical Imaging*, Vol.8, No.2:pp 143–153, 1989.

[MUR87] MURRAY D.W., BUXTON H. Scene segmentation from visual motion using global optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.9, No.2:pp 220–228, March 1987.

[NEL88] NELSON R.C., ALOIMONOS J. Using flow field divergence for obstacle avoidance: towards qualitative vision. *Proc. 2nd International Conference on Computer Vision*, pp 188–196, Dec. 1988.

[PEL90] PELEG S. , ROM H. motion based segmentation. *Proc. 10th IEEE Conference on Pattern Recognition, Atlantic City*, pp 109–113, 1990.

[SCH86] SCHUNCK B.G. The image flow constraint equation. *Computer Vision, Graphics and Image Processing*, Vol.35:pp 20–46, 1986.

[SUB89] SUBBARAO M. Interpretation of image flow: a spatio-temporal approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.11, No.3:pp 266–278, 1989.

[SUB90] SUBBARAO M. Bounds on time-to-collision and rotational component from first-order derivatives of image flow. *Computer Vision, Graphics and Image Processing*, Vol.50:pp 329–341, 1990.

[THO84] THOMPSON W.B., BERZINS V.A., MUTCH K.M. Analyzing object motion based on optical flow. *Proc. 7th IEEE Conference on Pattern Recognition, Montreal*, pp 791–794, 1984.

[THO90] THOMPSON W.B., PONG T.-G. Detecting moving objects. *International Journal of Computer Vision*, Vol.4:pp 39–57, 1990.

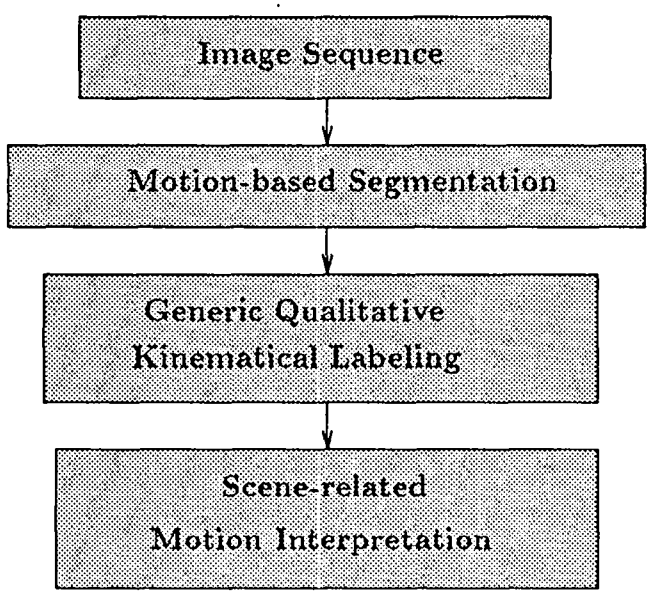


Figure 1: *Overview of the motion-identification scheme*

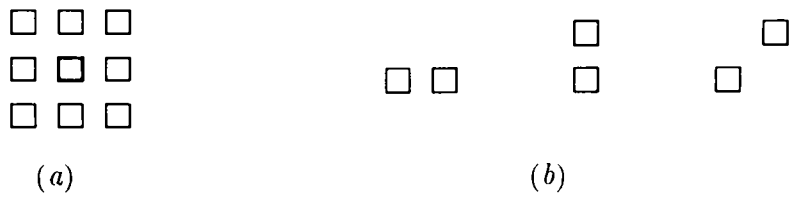


Figure 2: *(a) Second order neighborhood system, (b) binary cliques.*

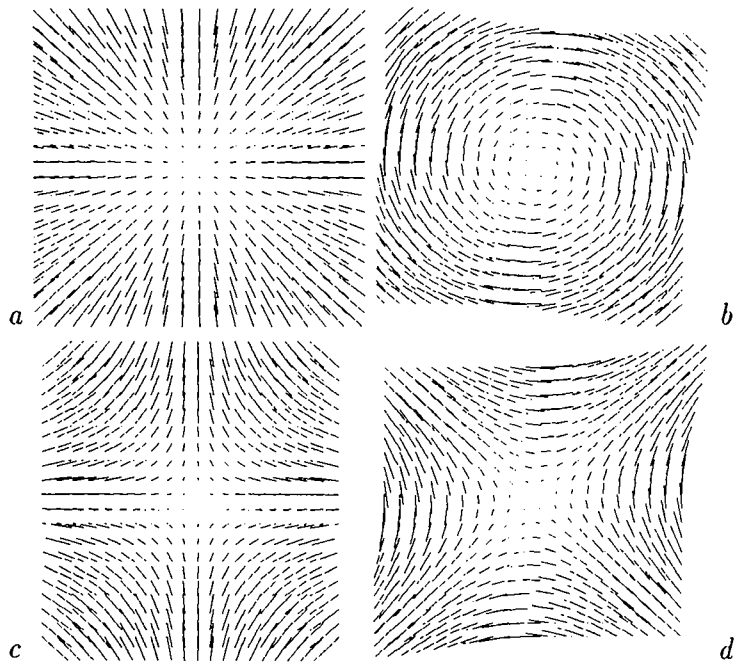


Figure 3: *(a) divergent field, (b) rotational field, (c) and (d) hyperbolic fields*

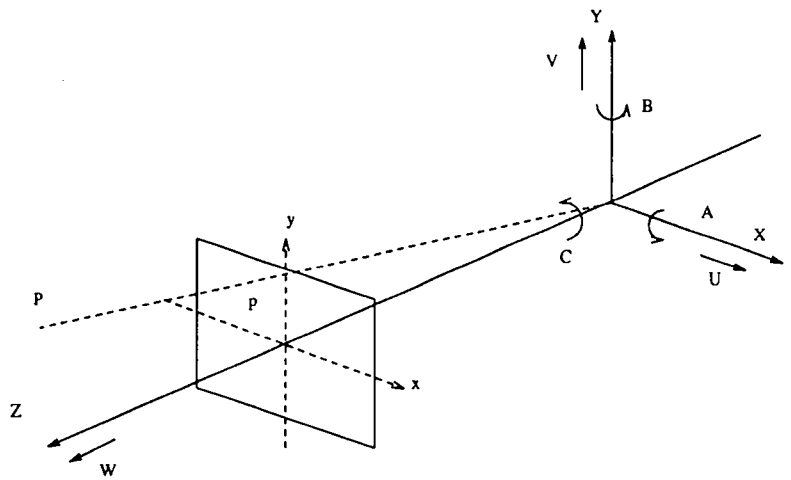


Figure 4: *coordinate system*

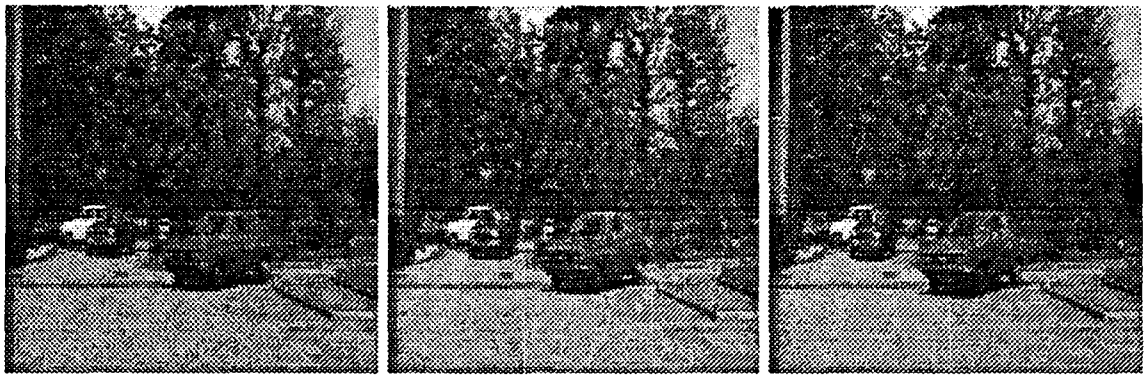


Figure 5: *Sequence 1: three images out of the sequence at time t_1, t_5 and t_9.*

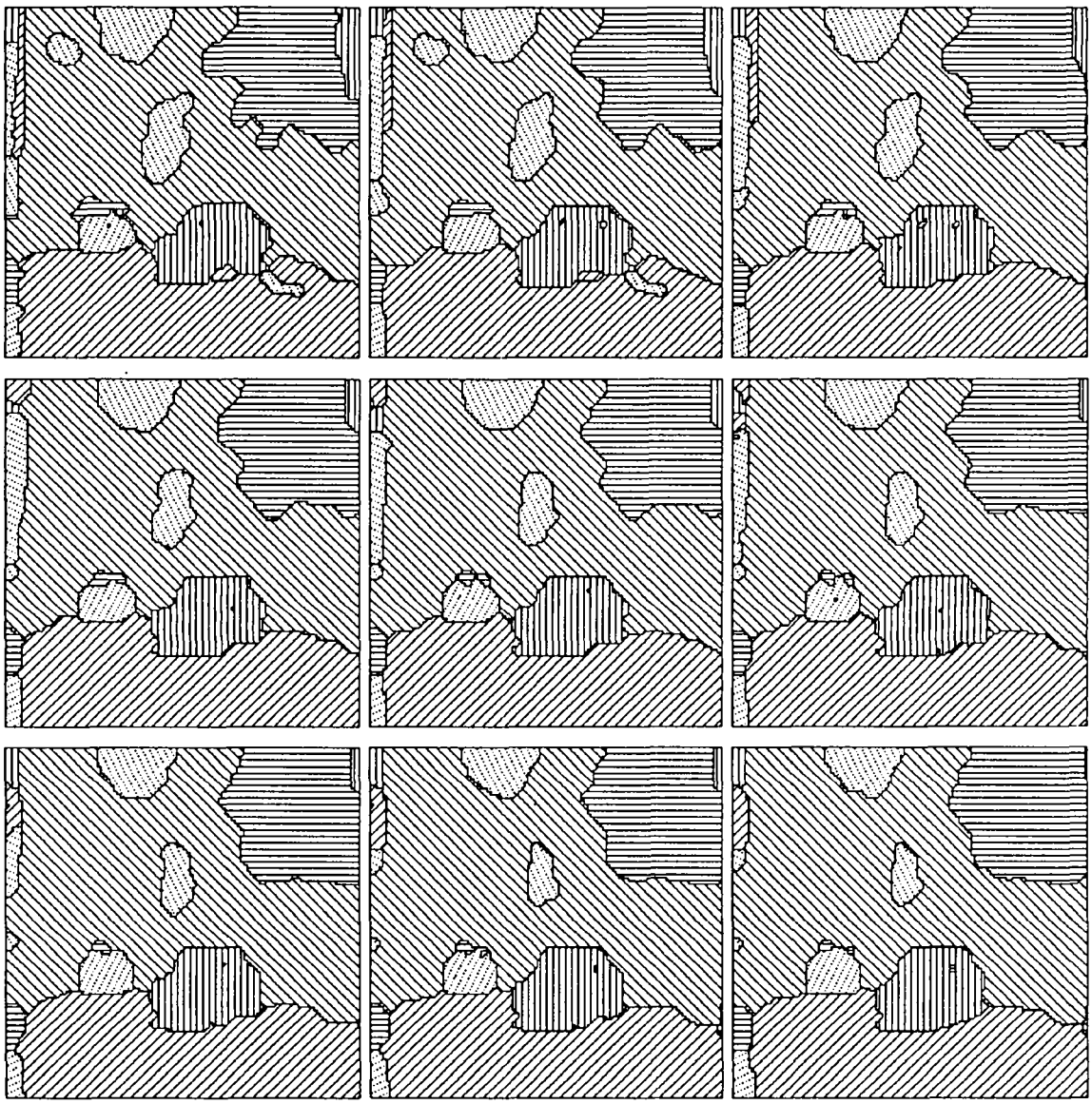


Figure 6: *Sequence 1: the motion-based segmentation maps from time t_1 to t_9.*

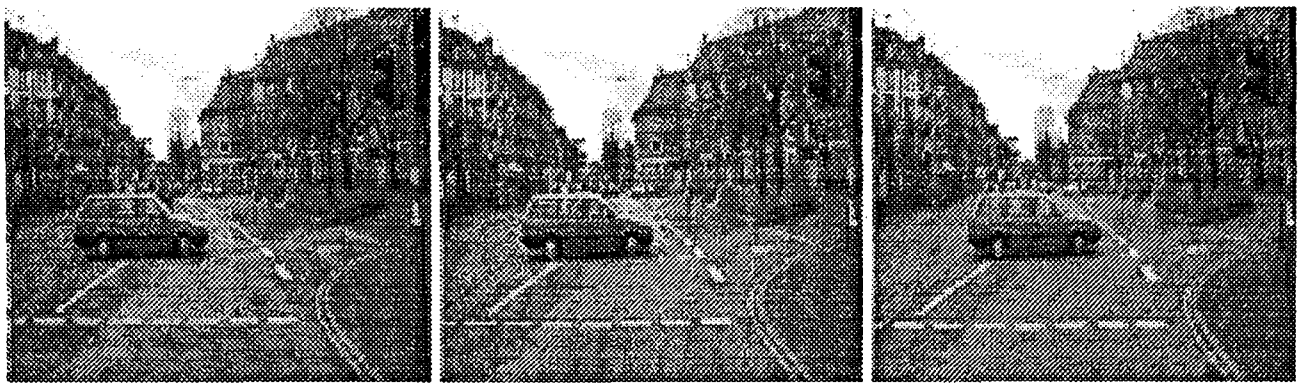


Figure 7: *Sequence 2: three images out of the sequence at time t_1, t_3 and t_5.*

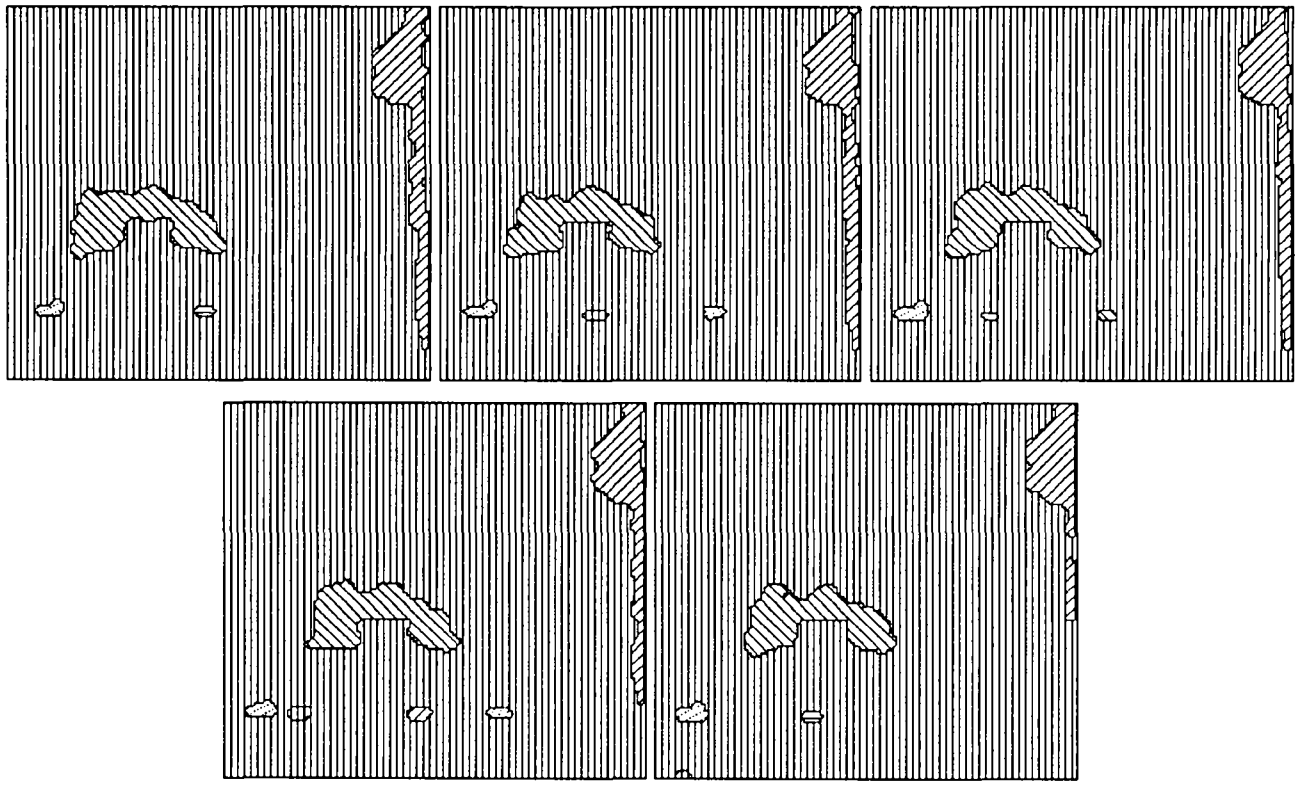


Figure 8: *Sequence 2: the motion-based segmentation maps from time t_1 to t_5.*

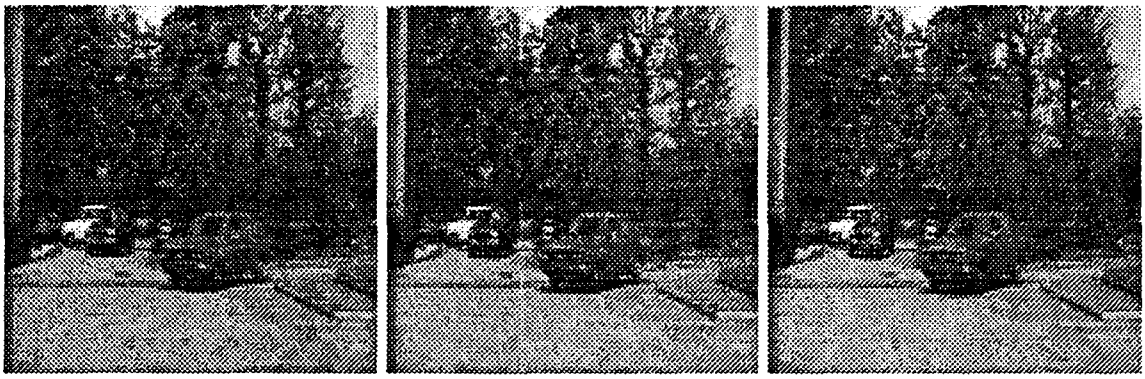


Figure 9: *Sequence 1: three images out of the sequence at time t_4, t_6 and t_8.*

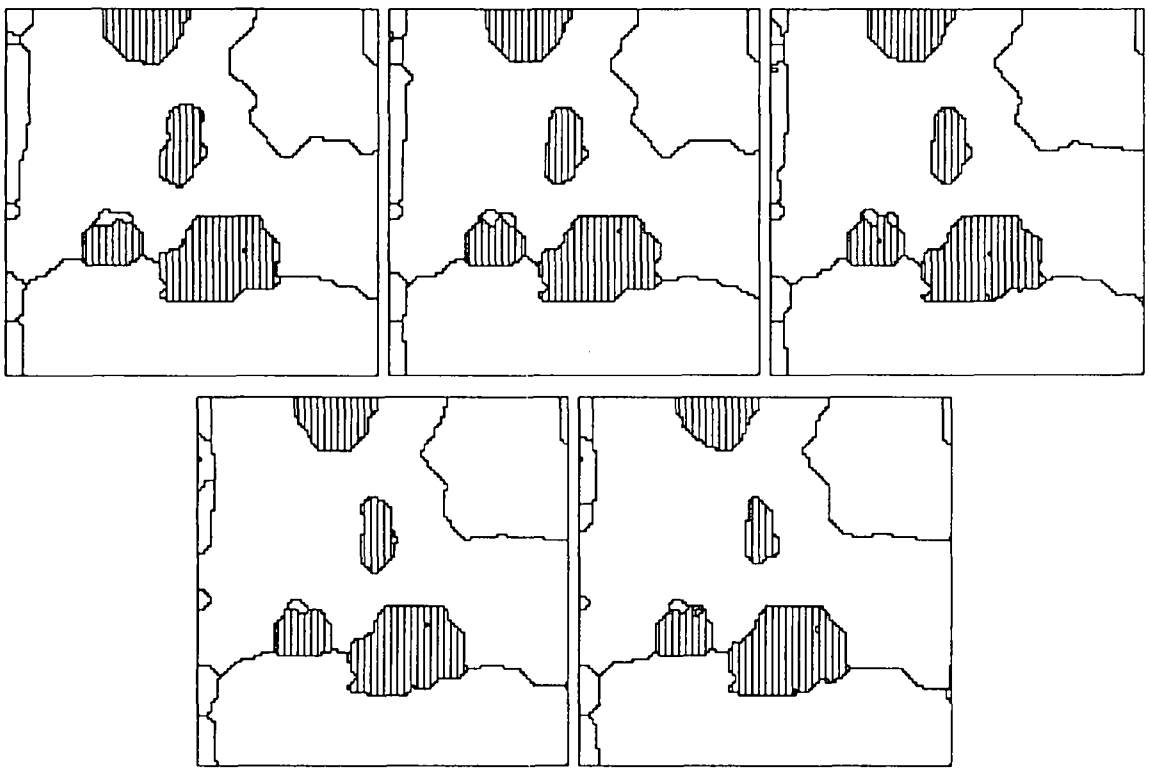
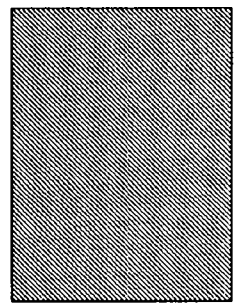
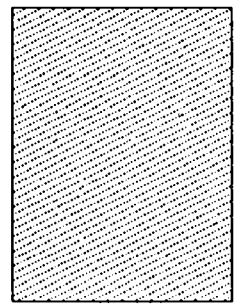


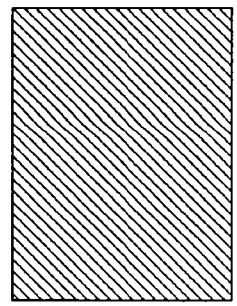
Figure 10: *Sequence 1: the scene-related motion interpretation maps from time t_4 to t_8.*



approaching



moving away



lateral motion

Figure 11: *The three classes for motion interpretation, and their corresponding texture.*

Figure 12: *Sequence 3: three images out of the sequence at at time t_1, t_5 and t_9.*

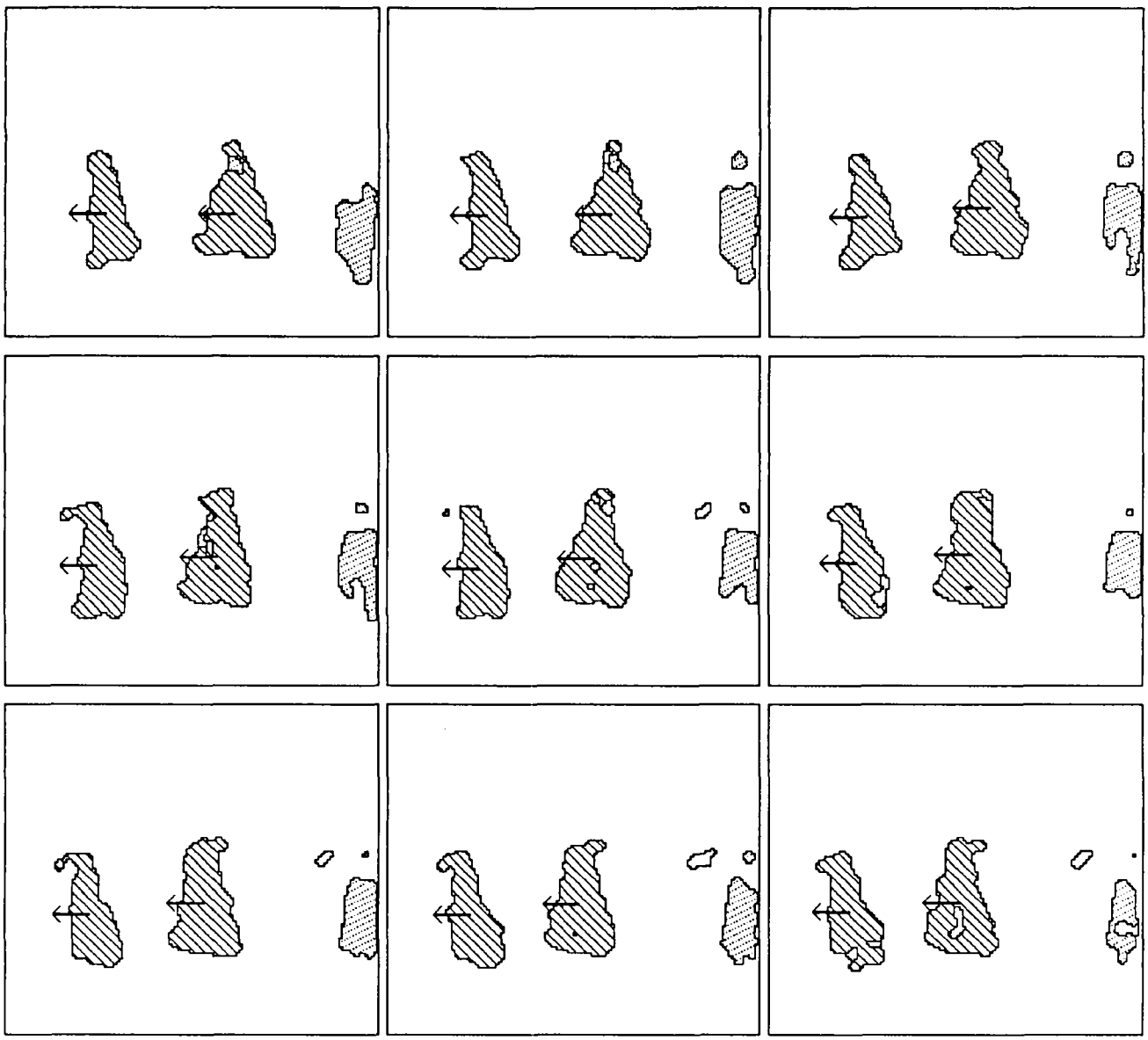


Figure 13: *Sequence 3: the scene-related motion interpretation maps from time t_1 to t_9.*

Figure 14: *Sequence 4: three images out of the sequence at time t_1, t_4 and t_7.*

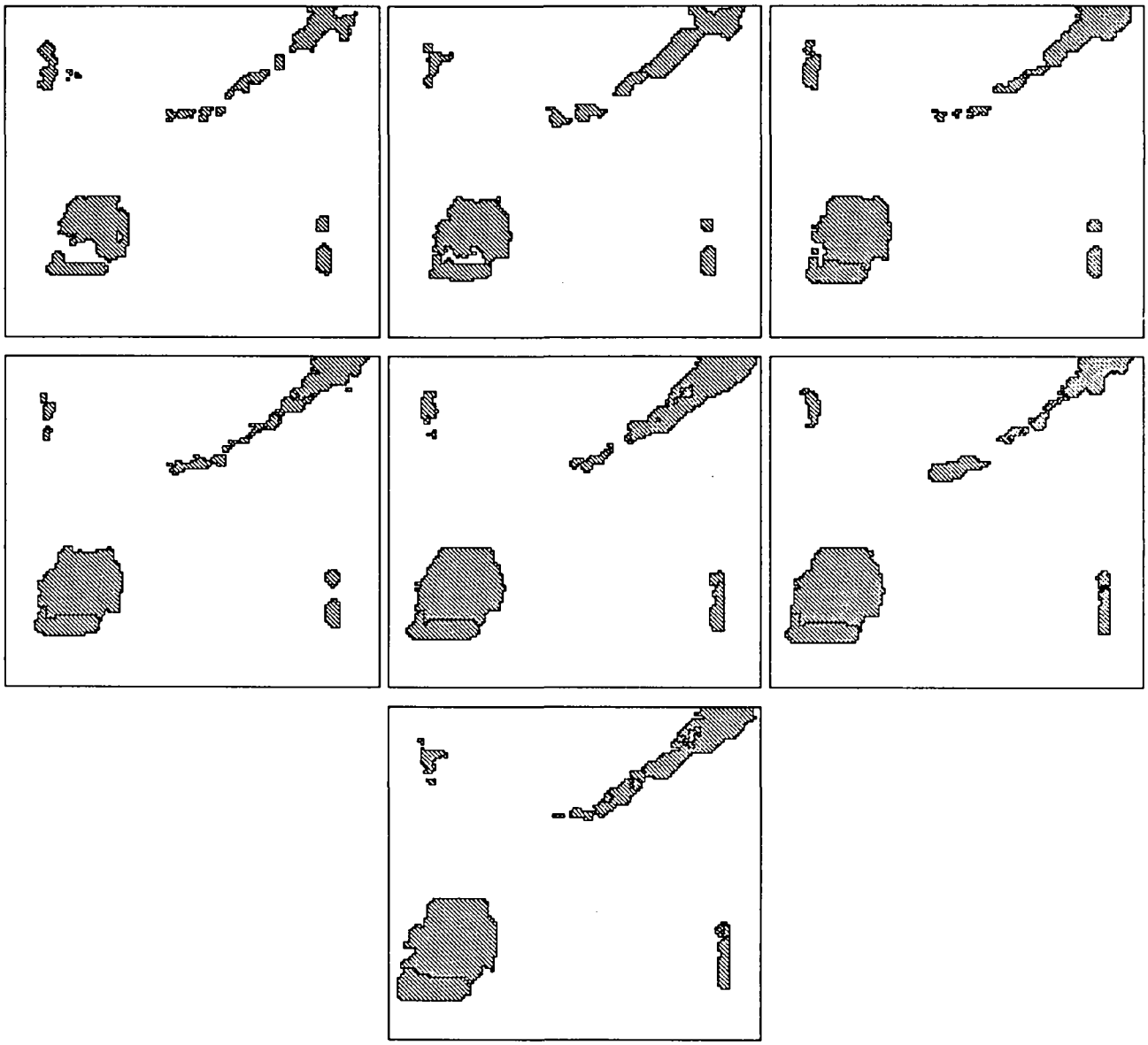


Figure 15: *Sequence 4: the scene-related motion interpretation maps from time t_1 to t_7.*

LISTE DES DERNIERES PUBLICATION INTERNES IRISA

PI 556 CONCEPTION ET INTEGRATION D'UN CORRELATEUR SYSTOLIQUE
Catherine DEZAN, Eric GAUTRIN, Patrice QUINTON
Novembre 1990, 16 Pages.

PI 557 VARIATIONAL APPROACH OF A MAGNETIC SHAPING PROBLEM
Michel CROUZEIX
Novembre 1990, 14 Pages.

PI 558 THE DAVIDSON METHOD
Michel CROUZEIX, Bernard PHILIPPE et Miloud SADKANE
Novembre 1990, 22 Pages.

PI 559 A DISTRIBUTED SOLUTION TO THE k-OUT OF-M RESOURCES ALLOCATION PROBLEM
Michel RAYNAL
Novembre 1990, 18 Pages.

PI 560 A SIMPLE TAXONOMY FOR DISTRIBUTED MUTUAL EXCLUSION ALGORITHMS
Michel RAYNAL
Novembre 1990.

PI 561 MULTIMODAL ESTIMATION OF DISCONTINOUS OPTICAL FLOW USING MARKOV RANDOM FIELDS
Fabrice HEITZ, Patrick BOUTHEMY
Novembre 1990, 50 Pages.

PI 562 EFFICIENT GLOBALCOMPUTATIONS ON A PROCESSOR NETWORK WITH PROGRAMMABLE LOGIC
J M. FILLOQUE, E. GAUTRIN, B POTTIER
Novembre 1990, 14 pages.

PI 563 EQUATIONAL SETS OF TREE-VECTORS
Anne GRAZON, Jean-Claude RAOULT
Novembre 1990, 20 Pages.

PI 564 MULTIFRAME-BASED IDENTIFICATION OF MOBILE COMPONENTS OF A SCENE WITH A MOVING CAMERA
Edouard FRANCOIS, Patrick BOUTHEMY
Décembre 1990, 30 pages.

PI 565 NAIVE RESERVE CAN BE LINEAR
Pascal BRISSET, Olivier RIDOUX
Novembre 1990, 18 pages.

ISSN 0249-6399